# CERTIFICATE OF VERIFICATION

I, Shiro TERASAKI

Japanese Patent Attorney

Registered No. 100092657

of SOEI PATENT AND LAW FIRM

Ginza First Bldg.

10-6 Ginza 1-chome

Chuo-ku, Tokyo 104-0061

Japan

state that the attached documents are a true and complete translation to the best of my knowledge of Japanese Patent Application No.2002-291610.

Dated this 19th day of February, 2008

Signature of translator:____________________________

Shiro TERASAKI

PATENT OFFICE

JAPANESE GOVERNMENT

This is to certify that the annexed is a true copy of the following application as filed with this Office.

Date of Application: October 3, 2002

Application Number: Japanese Patent Application No.2002-291610

Commissioner,

Patent Office (Seal)

(Document Name)Patent Application

(Reference Number) 14-0405

(Directly) Commissioner of the Patent Office

(IPC) H04N 1/41

(Inventor)

(Residence or Address)

c/o Intellectual Property Department, NTT DoCoMo, Inc., 11-1, Nagatacho 2-chome, Chiyoda-ku, Tokyo

(Name) Satoru ADACHI

(Inventor)

(Residence or Address)

c/o Intellectual Property Department, NTT DoCoMo, Inc., 11-1, Nagatacho 2-chome, Chiyoda-ku, Tokyo

(Name) Minoru ETOH

(Inventor)

(Residence or Address)

c/o Intellectual Property Department, NTT DoCoMo, Inc., 11-1, Nagatacho 2-chome, Chiyoda-ku, Tokyo

(Name) Sadaatsu KATO

(Inventor)

(Residence or Address)

c/o Intellectual Property Department, NTT DoCoMo, Inc., 11-1, Nagatacho 2-chome, Chiyoda-ku, Tokyo

(Name) Thiow Keng TAN

(Applicant)

(Identification Number) 392026693

(Name) NTT DoCoMo, Inc.

(Attorney)

(Identification Number) 100088155

(Patent Attorney)

(Name) Yoshiki HASEGAWA

(Attorney)

(Identification Number) 100092657

(Patent Attorney)

(Name) Shiro TERASAKI

(Attorney)

(Identification Number) 100114270

(Patent Attorney)

(Name) Tomoya KUROKAWA

(Attorney)

(Identification Number) 100108213

(Patent Attorney)

(Name) Toyotaka ABE

(Attorney)

(Identification Number) 100113549

(Patent Attorney)

(Name) Mamoru SUZUKI

(Official Fee)

(Pre-Paid Master Note Number) 014708

(Amount to be paid) 21000

(Lists of the Article to be presented)

(Name of Article) Specification 1

(Name of Article) Drawings 1

(Name of Article) Abstract 1

(Proof Reading) Required

[Document Name] Specification

[Title of the Invention] VIDEO ENCODING METHOD, VIDEO DECODING METHOD, VIDEO ENCODING APPARATUS, VIDEO DECODING APPARATUS, VIDEO PROCESSING SYSTEM, VIDEO ENCODING PROGRAM, AND VIDEO DECODING PROGRAM

[Claims]

[Claim 1] A video encoding method of implementing backward interframe prediction from a temporally subsequent frame, said video encoding method comprising:

outputting a maximum delay time that can be made by backward prediction.

[Claim 2] The video encoding method according to Claim 1, wherein said maximum delay time is defined as a time difference between an occurrence time of a frame to be subjected to backward interframe prediction, and an occurrence time of a temporally last subsequent frame that can be used as a reference frame in backward prediction.

[Claim 3] The video encoding method according to Claim 1 or 2, wherein the maximum delay time is outputted as information to be applied to entire encoded data.

[Claim 4] The video encoding method according to Claim 1 or 2, wherein the maximum delay time is outputted as information to be applied to each frame.

[Claim 5] The video encoding method according to Claim 1 or 2, wherein the maximum delay time is optionally outputted as information to be applied to a frame for which the maximum delay time

is transmitted and to each temporally subsequent frame after said frame.

[Claim 6] A video decoding method of implementing backward interframe prediction from a temporally subsequent frame, said video decoding method comprising:

effecting input of a maximum delay time that can be made by backward prediction.

[Claim 7] The video decoding method according to Claim 6, wherein the maximum delay time is defined as a time difference between a decoding time of a frame without delay due to backward interframe prediction and without reversal of orders of decoding times and output times with respect to any other frame, and a decoded image output time correlated with said frame, and

a reference for decoded image output times thereafter is set based on the maximum delay time.

[Claim 8] The video decoding method according to Claim 6 or 7, wherein the maximum delay time is entered as information to be applied to entire encoded data.

[Claim 9] The video decoding method according to Claim 6 or 7, wherein the maximum delay time is entered as information to be applied to each frame.

[Claim 10] The video decoding method according to Claim 6 or 7, wherein the maximum delay time is optionally entered as information to be applied to a frame for which the maximum delay time is transmitted and to each temporally subsequent frame after said frame.

[Claim 11] A video encoding apparatus for implementing backward interframe prediction from a temporally subsequent frame,

said video encoding apparatus being configured to:

output a maximum delay time that is incurred by backward prediction.

[Claim 12] The video encoding apparatus according to Claim 11, wherein said maximum delay time is defined as a time difference between an occurrence time of a frame to be subjected to backward interframe prediction, and an occurrence time of a temporally last subsequent frame that can be used as a reference frame in backward prediction.

[Claim 13] The video encoding apparatus according to Claim 11 or 12, wherein the maximum delay time is outputted as information to be applied to entire encoded data.

[Claim 14] The video encoding apparatus according to Claim 11 or 12, wherein the maximum delay time is outputted as information to be applied to each frame.

[Claim 15] The video encoding apparatus according to Claim 11 or 12, wherein the maximum delay time is optionally outputted as information to be applied to a frame for which the maximum delay time is transmitted and to each temporally subsequent frame after said frame.

[Claim 16] A video decoding apparatus for implementing backward interframe prediction from a temporally subsequent frame, said video decoding apparatus being configured to:

effect input of a maximum delay time that is incurred by backward prediction.

[Claim 17] The video decoding apparatus according to Claim 16, wherein the maximum delay time is defined as a time difference

between a decoding time of a frame without delay due to backward interframe prediction and without reversal of orders of decoding times and output times with respect to any other frame, and a decoded image output time correlated with said frame, and

a reference for decoded image output times thereafter is set based on the maximum delay time.

[Claim 18] The video decoding apparatus according to Claim 16 or 17, wherein the maximum delay time is entered as information to be applied to entire encoded data.

[Claim 19] The video decoding apparatus according to Claim 16 or 17, wherein the maximum delay time is entered as information to be applied to each frame.

[Claim 20] The video decoding apparatus according to Claim 16 or 17, wherein the maximum delay time is optionally entered as information to be applied to a frame for which the maximum delay time is transmitted and to each temporally subsequent frame after said frame.

[Claim 21] A video processing system comprising a video encoding apparatus and a video decoding apparatus, wherein

the encoding apparatus is the video encoding apparatus according to Claim 11, and

the decoding apparatus is the video decoding apparatus according to Claim 16.

[Claim 22] A video encoding program for letting a computer to execute video encoding of implementing backward interframe prediction from a temporally subsequent frame, said video encoding program letting the computer to execute:

a process of outputting a maximum delay time that is incurred by backward prediction.

[Claim 23] The video encoding program according to Claim 22, wherein said maximum delay time is defined as a time difference between an occurrence time of a frame to be subjected to backward interframe prediction, and an occurrence time of a temporally last subsequent frame that can be used as a reference frame in backward prediction.

[Claim 24] The video encoding program according to Claim 22 or 23, wherein the maximum delay time is outputted as information to be applied to entire encoded data.

[Claim 25] The video encoding program according to Claim 22 or 23, wherein the maximum delay time is outputted as information to be applied to each frame.

[Claim 26] The video encoding program according to Claim 22 or 23, wherein the maximum delay time is optionally outputted as information to be applied to a frame for which the maximum delay time is transmitted and to each temporally subsequent frame after said frame.

[Claim 27] A video decoding program for letting a computer to execute video decoding of implementing backward interframe prediction from a temporally subsequent frame, said video decoding program letting the computer to execute:

a process of effecting input of a maximum delay time that can be made by backward prediction.

[Claim 28] The video decoding program according to Claim 27, wherein the maximum delay time is defined as a time difference

between a decoding time of a frame without delay due to backward interframe prediction and without reversal of orders of decoding times and output times with respect to any other frame, and a decoded image output time correlated with said frame, and

said video decoding program letting the computer to execute a process of setting a reference for decoded image output times thereafter based on the maximum delay time.

[Claim 29] The video decoding program according to Claim 27 or 28, wherein the maximum delay time is entered as information to be applied to entire encoded data.

[Claim 30] The video decoding program according to Claim 27 or 28, wherein the maximum delay time is entered as information to be applied to each frame.

[Claim 31] The video decoding program according to Claim 27 or 28, wherein the maximum delay time is optionally entered as information to be applied to a frame for which the maximum delay time is transmitted and to each temporally subsequent frame after said frame.

[Detailed Description of the Invention]

[0001]

[Technical Field to which the Invention Pertains]

The present invention relates to a video encoding method, a video decoding method, a video encoding apparatus, a video decoding apparatus, a video processing system, a video encoding program, and a video decoding program.

[0002]

[Prior Art]

Video signal encoding techniques are used for transmission and storage-regeneration of video signals. The well-known techniques include, for example, the international standard video coding methods such as ITU-T Recommendation H.263 (hereinafter referred to as H.263), ISO/IEC International Standard 14496-2 (MPEG-4 Visual, hereinafter referred to as MPEG-4), and so on. Another known newer encoding method is a video coding method scheduled for joint international standardization by ITU-T and ISO/IEC; ITU-T Recommendation H.264 and ISO/IEC International Standard 14496-10 (Joint Final Committee Draft of Joint Video Specification, Non-Patent Document 1 (ftp://ftp.imtc-files.org/jvt-experts/2002_07_Klagenfurt/JVT-D157.zip), hereinafter referred to as H.26L). The general coding techniques used in these video encoding methods are described, for example, in Non-Patent Document 2 ("Basic Technologies on International Image Coding Standards" co-authored by Fumitaka Ono and Hiroshi Watanabe).

[0003]

Since a motion video signal consists of a series of images (frames) varying little by little with time, it is common practice in these video coding methods to implement interframe prediction between a frame retrieved as a target for encoding (current frame) and another frame (reference frame) and thereby reduce temporal redundancy in the video signal. In this case, where the interframe prediction is carried out between the current frame and a reference frame less different from the current frame, the redundancy can be reduced more and encoding

efficiency can be increased.

[0004]

For this reason, as shown in Fig. 4, the reference frame for the current frame A1 can be either a temporally previous frame A0 or a temporally subsequent frame A2 with respect to the current frame A1. The prediction with the previous frame is referred to as forward prediction, while the prediction with the subsequent frame as backward prediction. Bidirectional prediction is defined as a prediction in which one is arbitrarily selected out of the two prediction methods, or as a prediction in which both methods are used simultaneously.

[0005]

In general, with use of such bidirectional prediction, as in the example shown in Fig. 4, a temporally previous frame as a reference frame for forward prediction and a temporally subsequent frame as a reference frame for backward prediction each are preliminarily stored prior to the current frame.

[0006]

Fig. 5 is a figure including diagrams showing (a) decoding and (b) output of the frames in the case of the bidirectional prediction shown in Fig. 4. For example, in the decoding of MPEG-4, where the current frame A1 is decoded by bidirectional interframe prediction, frame A0 being one temporally previous frame and frame A2 being one temporally subsequent frame with respect to the current frame A1 are first decoded as frames decoded by intraframe prediction without use of interframe prediction or as frames decoded by forward interframe prediction, prior to decoding of the current frame A1, and they are

retained as reference frames. Thereafter, the current frame A1 is decoded by bidirectional prediction using these two frames A0, A2 thus retained (Fig. 5(a)).

[0007]

In this case, therefore, the order of decoding times of the temporally subsequent reference frame A2 and the current frame A1 is reverse to the order of output times of their respective decoded images. Each of these frames A0, A1, and A2 is attached with output time information 0, 1, or 2, and thus the temporal sequence of the frames can be known according to this information. For this reason, the decoded images are outputted in the right order (Fig. 5(b)).

[0008]

Some of the recent video coding methods permit the foregoing interframe prediction to be carried out using multiple reference frames, instead of one reference frame in the forward direction and one reference frame in the backward direction, so as to enable prediction from a frame with a smaller change from the current frame, as shown in Fig. 6. Fig. 6 shows an example using two temporally previous frames B0, B1 and two temporally subsequent frames B3, B4 with respect to the current frame B2, as reference frames for the current frame B2.

[0009]

Fig. 7 is a figure including diagrams showing (a) decoding and (b) output of the frames in the case of the bidirectional prediction shown in Fig. 6. For example, in the decoding of H.26L, a plurality of reference frames can be retained within a range up to a predetermined upper bound of the number of reference frames and, on the occasion of

carrying out interframe prediction, an optimal reference frame is arbitrarily designated out of them. In this case, where the current frame B2 is decoded as a bidirectionally predicted frame, the reference frames are first decoded prior to the decoding of the current frame B2; the reference frames include a plurality of temporally previous frames (e.g., two frames B0, B1) and a plurality of temporally subsequent frames (e.g., two frames B3, B4) with respect to the current frame B2, which are decoded and retained as reference frames. The current frame B2 can be predicted from a frame arbitrarily designated as the one used for prediction out of those frames B0, B1, B3, and B4 (Fig. 7(a)).

[0010]

In this case, therefore, the order of decoding times of the temporally subsequent reference frames B3, B4 and the current frame B2 becomes reverse to the order of their respective output times. Each of these frames B0-B4 is attached with output time information or output order information 0-4, and the temporal sequence of the frames can be known according to this information. For this reason, the decoded images are outputted in the right order (Fig. 7(b)).

[0011]

For carrying out the decoding by the backward prediction using temporally subsequent frames as predictive frames, it is necessary to satisfy the condition that the decoding of the temporally subsequent frames is completed prior to the decoding of the current frame so as to be available as predictive frames. In this case, a delay is incurred before the decoded image of the current frame becomes available, as compared with a frame to which the backward prediction is not applied.

[0012]

This will be specifically described below with reference to Fig. 8. Fig. 8 corresponds to the example shown in Figs. 4 and 5. First, encoded data of each frame A0-A2 is decoded in an order necessary for execution of interframe prediction, and it is assumed that intervals of the frames are constant time intervals according to a frame rate and that the time necessary for the decoding operation is negligible for each frame A0-A2, regardless of whether the interframe prediction is applied and regardless of the directions of interframe prediction (Fig. 8(a)). In practice, the decoding intervals of the frames A0-A2 do not have to be constant and can change depending upon such factors as variation in encoding bits of the frames A0-A2 or the like; however, they can be assumed to be constant on average. The time necessary for the decoding operation is not zero, either, but it will raise no significant problem in the description hereinafter if the difference thereof is not so large among the frames A0-A2.

[0013]

It is supposed herein that a time when a decoded image of frame A0 without delay due to backward prediction and without reversal of the orders of decoding times and output times with respect to any other frame (a frame without delay and without reversal will be referred to hereinafter as a backward-prediction-nonassociated frame) is obtained, is defined as an output time correlated with the decoded image, and the decoded image is outputted at the output time. Supposing the subsequent frame is the backward predicted frame A1, the decoded image thereof will be decoded after the temporally subsequent frame

A2, and a delay is thus made before the decoded image is obtained.

[0014]

For this reason, if the time when the decoded image is obtained for the backward-prediction-nonassociated frame A0 is defined as a reference of output time, the decoded image of the backward predicted frame A1 is not obtained by the output time correlated therewith (Fig. 8(b)). Namely, an output time interval between the decoded image of the backward-prediction-nonassociated frame A0 and the decoded image of the backward predicted frame A1 becomes longer by the delay time necessary for execution of backward prediction than the original interval, which leads to unnatural video output.

[0015]

Therefore, in the case where the backward interframe prediction is applied in video coding, as shown in Fig. 8(c), it is necessary to preliminarily delay the output time of the decoded image of the backward-prediction-nonassociated frame A0 by the delay time necessary for execution of the backward prediction as well so as to be able to correctly handle the output time interval to the backward predicted frame A1.

[0016]

Conventionally, the backward interframe prediction was applied to video encoding under the conditions that encoding was carried out at a high bit rate and the fixed frame rate of 30 frames/second equal to that of TV broadcast signals was always used, like TV broadcasting or accumulation thereof, because backward interframe prediction brings about more options for prediction and hence increase of computational

complexity so as to make implementation thereof difficult on simple equipment and because the increase of delay time was not desired in real-time communication involving bidirectional interlocution like video conferences.

[0017]

In this case, for example, as in MPEG-4, where the use of one temporally subsequent frame as a reference frame for backward prediction, the delay time necessitated in execution of the backward prediction is constant. For example, where the frame rate is 30 frames/second as described above, the delay time is a time interval of each frame, i.e., 1/30 second. Accordingly, the time by which the output time of the decoded image of the backward-prediction-nonassociated frame should be delayed, can be equally set to 1/30 second.

[0018]

[Non-Patent Document 1]

(searched on October 1, 2002 (Hei. 14)), Internet <ftp address : ftp://ftp.imtc-files.org/jvt-experts/2002_07_Klagenfurt/JVT-D157.zip>

[Non-Patent Document 2]

"Basic Technologies on International Image Coding Standards" co-authored by Fumitaka Ono and Hiroshi Watanabe, and published on March 20, 1998 by CORONA PUBLISHING CO., LTD.

[0019]

[Problem to be Solved by the Invention]

In recent years, however, following the improvement in computer performance and progress in diversification of video services,

delay is tolerable in video delivery through the Internet and mobile communications, and there is increased use of video coding requiring encoding at low bit rates. For implementing the encoding at low bit rates, frame rates smaller than 30 frames/second are applied, or variable frame rates are used to dynamically change the frame rate in order to control the encoding bit rate.

[0020]

In such video coding, where the aforementioned backward prediction is applied in order to increase the encoding efficiency more, the delay time due to the backward prediction is not always 1/30 second as used before. In the application of variable frame rates, the frame rates are not constant. For example, in the case where a small frame rate is used on a temporary basis, the time interval of each frame there becomes large, and thus the time by which the output time of the decoded image of the backward-prediction-nonassociated frame should be delayed is not uniquely determined. For this reason, it becomes infeasible to correctly handle the output time interval between the decoded image of the backward-prediction-nonassociated frame and the decoded image of the backward predicted frame.

[0021]

In this case, there is such potential means that a large permissible delay time is preliminarily allowed for the backward prediction and that the output time of the decoded image of the backward-prediction-nonassociated frame is always delayed by this delay time, thereby correctly handling the output time interval relative to the decoded image of the backward predicted frame. In this case,

however, the large delay is always added to the output time of the decoded image, regardless of the delay time in the practical backward prediction.

[0022]

When multiple reference frames are used in the backward prediction as in H.26L, the decoding of all the reference frames being temporally subsequent frames must be completed prior to the decoding of the current frame. This further increases the delay time necessary for execution of the backward prediction.

[0023]

In this case, since the number of reference frames used in the backward prediction is uniquely determined as a number of temporally subsequent frames to the current frame, which were decoded prior to the current frame, the number of reference frames can be optionally changed within the range up to the predetermined upper bound of the maximum number of reference frames.

[0024]

For example, supposing the upper bound of the number of reference frames is 4, the number of reference frames used in the backward prediction may be 2 as shown in Fig. 6, or 1 as shown in Fig. 9(a), or 3 as shown in Fig. 9(b). Since the number of reference frames can be changed in this way, the delay time necessary for execution of the backward prediction can vary largely. This leads to failure in correctly handling the output time interval between the decoded image of the backward-prediction-nonassociated frame and the decoded image of the backward predicted frame.

[0025]

At this time, since the maximum number of reference frames that can be used in the backward prediction does not exceed the upper bound of the number of reference frames, the delay time according to the upper bound of the number of reference frames is a maximum delay time that can be made in execution of the backward prediction. Therefore, if the output time of the decoded image of the backward-prediction-nonassociated frame is always delayed by this delay time, the output time interval relative to the decoded image of the backward predicted frame can be correctly handled.

[0026]

In this case, however, a large delay is always added to the output time of the decoded image, regardless of the number of reference frames actually used for the backward predicted frame. In the application of variable frame rates as described above, while the maximum number of reference frames can be uniquely determined, the maximum delay time cannot be uniquely determined.

[0027]

In the application of the backward prediction to the video coding heretofore, it was infeasible to uniquely determine the delay time necessary for execution of the backward prediction, except for the case where use of a fixed frame rate was clear. This resulted in failure in correctly handling the output time interval between the decoded image of the backward-prediction-nonassociated frame and the decoded image of the backward predicted frame, thus posing the problem that the video output became unnatural.

[0028]

In the case where multiple reference frames are used in the backward prediction, the number of reference frames can also be changed, so as to possibly vary the delay time. Therefore, there is the problem of the failure in correctly handling the time interval between the decoded image of the backward-prediction-nonassociated frame and the decoded image of the backward predicted frame. In the case where the maximum delay time is always assumed in order to cope with this problem, there arises the problem that the large delay is always added to the output time of the decoded image.

[0029]

The present invention has been accomplished in order to solve the above problems, and an object of the invention is to provide a video encoding method, a video decoding method, a video encoding apparatus, a video decoding apparatus, a video processing system, a video encoding program, and a video decoding program capable of achieving output of decoded images at appropriate time intervals when employing backward interframe prediction.

[0030]

[Means for Solving the Problem]

In order to achieve the above object, a video encoding method according to the present invention is a video encoding method of implementing backward interframe prediction from a temporally subsequent frame, the video encoding method comprising: outputting a maximum delay time that is incurred by backward prediction.

[0031]

Likewise, a video encoding apparatus according to the present invention is a video encoding apparatus for implementing backward interframe prediction from a temporally subsequent frame, the video encoding apparatus being configured to: output a maximum delay time that is incurred by backward prediction.

[0032]

In the video encoding method and apparatus according to the present invention, as described above, on the occasion of encoding a moving picture consisting of a series of frames and outputting encoded data, the maximum delay time due to the backward prediction is outputted in addition to the encoded data. This enables achievement of output of decoded images at appropriate time intervals when employing the backward interframe prediction.

[0033]

A video encoding program according to the present invention is a video encoding program for letting a computer to execute video encoding of implementing backward interframe prediction from a temporally subsequent frame, the video encoding program letting the computer to execute: a process of outputting a maximum delay time that is incurred by backward prediction.

[0034]

In the video encoding program according to the present invention, as described above, on the occasion of encoding a moving picture and outputting encoded data thereof, the computer is made to execute the process of outputting the maximum delay time, in addition to the encoded data. This enables achievement of output of decoded

images at appropriate time intervals when employing the backward interframe prediction.

[0035]

A video decoding method according to the present invention is a video decoding method of implementing backward interframe prediction from a temporally subsequent frame, the video decoding method comprising: effecting input of a maximum delay time that can be made by backward prediction.

[0036]

Likewise, a video decoding apparatus according to the present invention is a video decoding apparatus for implementing backward interframe prediction from a temporally subsequent frame, the video decoding apparatus being configured to: effect input of a maximum delay time that is incurred by backward prediction.

[0037]

In the video decoding method and apparatus according to the present invention, as described above, on the occasion of decoding input encoded data to generate a moving picture, the maximum delay time due to the backward prediction is entered in addition to the encoded data. This enables achievement of output of decoded images at appropriate time intervals when employing the backward interframe prediction.

[0038]

A video decoding program according to the present invention is a video decoding program for letting a computer to execute video decoding of implementing backward interframe prediction from a

temporally subsequent frame, the video decoding program letting the computer to execute: a process of effecting input of a maximum delay time that is incurred by backward prediction.

[0039]

In the video decoding program according to the present invention, as described above, on the occasion of decoding encoded data to generate a moving picture, the computer is made to execute the process of effecting the input of the maximum delay time, in addition to the encoded data. This enables achievement of output of decoded images at appropriate time intervals when employing the backward interframe prediction.

[0040]

Concerning the maximum delay time outputted in the video encoding method, encoding apparatus, and encoding program, it is preferable to define the maximum delay time as a time difference between an occurrence time of a frame to be subjected to backward interframe prediction and an occurrence time of a temporally last subsequent frame that can be used as a reference frame in backward prediction.

[0041]

Concerning application of the maximum delay time, the maximum delay time may be outputted as information to be applied to the entire encoded data. In another embodiment, the maximum delay time may be outputted as information to be applied to each frame. In still another embodiment, the maximum delay time may be optionally outputted as information to be applied to a frame for which the

maximum delay time is indicated and to each temporally subsequent frame after the foregoing frame.

[0042]

Concerning the maximum delay time entered in the video decoding method, decoding apparatus, and decoding program, it is preferable to define the maximum delay time as a time difference between a decoding time of a frame without delay due to backward interframe prediction and without reversal of orders of decoding times and output times with respect to any other frame, and a decoded image output time correlated with the foregoing frame. In another embodiment, furthermore, it is preferable to set a reference for decoded image output times thereafter on the basis of the maximum delay time.

[0043]

Concerning application of the maximum delay time, the maximum delay time may be entered as information to be applied to the entire encoded data. In another embodiment, the maximum delay time may be entered as information to be applied to each frame. In still another embodiment, the maximum delay time may be optionally entered as information to be applied to a frame for which the maximum delay time is indicated and to each temporally subsequent frame after the foregoing frame.

[0044]

A video processing system according to the present invention is a video processing system comprising a video encoding apparatus and a video decoding apparatus, wherein the encoding apparatus is the video encoding apparatus described above and wherein the decoding

apparatus is the video decoding apparatus described above.

[0045]

As described above, the video processing system according to the present invention is constructed using the video encoding apparatus and the video decoding apparatus for effecting output and input of the maximum delay time due to the backward prediction. This substantializes the video processing system capable of achieving output of decoded images at appropriate time intervals when employing the backward interframe prediction.

[0046]

[Embodiments of the Invention]

The preferred embodiments of the video encoding method, video decoding method, video encoding apparatus, video decoding apparatus, video processing system, video encoding program, and video decoding program according to the present invention will be described below in detail with reference to the drawings. The same elements will be denoted by the same reference symbols throughout the description of the drawings, without redundant description thereof.

[0047]

First, the encoding and decoding of moving picture in the present invention will be schematically described. Fig. 1 is a block diagram showing the schematic structure of the video encoding apparatus, video decoding apparatus, and video processing system according to the present invention. The video processing system is comprised of video encoding apparatus 1 and video decoding apparatus 2. The video encoding apparatus 1, video decoding apparatus 2, and

video processing system will be described below together with the video encoding method and video decoding method executed therein.

[0048]

The video encoding apparatus 1 is a device configured to encode video data D0 consisting of a series of images (frames) and output encoded data D1, for transmission, for storage and regeneration of moving pictures. The video decoding apparatus 2 is a device configured to decode input encoded data D1 to generate decoded moving picture data D2 consisting of a series of frames. The video encoding apparatus 1 and the video decoding apparatus 2 are connected by a predetermined wired or wireless data transmission line, in order to transmit necessary data such as the encoded data D1 and others.

[0049]

In the encoding of the moving picture carried out in the video encoding apparatus 1, as described previously, the interframe prediction is carried out between a frame of video data D0 entered as a target for encoding, and another frame as a reference frame, thereby reducing the redundancy in the video data. In the video processing system shown in Fig. 1, the video encoding apparatus 1 carries out the backward interframe prediction from a temporally subsequent frame for interframe prediction. Furthermore, this video encoding apparatus 1 outputs the maximum delay time that is incurred by the backward prediction, in addition to the encoded data D1.

[0050]

In correspondence to such video encoding apparatus 1, the video decoding apparatus 2 is configured to effect input of the maximum

delay time that is incurred by the backward prediction, in addition to the encoded data D1 from the video encoding apparatus 1. Then the video decoding apparatus 2 decodes the encoded data D1 with reference to the input maximum delay time to generate the video data D2.

[0051]

By the video encoding apparatus 1 and video encoding method configured to output the maximum delay time, the video decoding apparatus 2 and video decoding method configured to effect input of the maximum delay time, and the video processing system equipped with those apparatus 1, 2, which are adapted for the backward interframe prediction as described above, it becomes feasible to achieve output of decoded images at appropriate time intervals in execution of the interframe prediction using the backward interframe prediction.

[0052]

Concerning the maximum delay time outputted in the video coding, for example, the maximum delay time can be defined as a time difference between an occurrence time of a frame to be subjected to the backward interframe prediction and an occurrence time of a temporally last subsequent frame that can be used as a reference frame for backward prediction.

[0053]

As for the maximum delay time entered in the video decoding, for example, the maximum delay time can be defined as a time difference between a decoding time of a frame without delay due to backward interframe prediction and without reversal of orders of decoding times and output times with respect to other frame and a

decoded image output time correlated with the pertinent frame. In this case, preferably, a reference for decoded image output times thereafter is set based on the maximum delay time.

[0054]

Application of the maximum delay time can be a method of applying it to entire encoded data or a method of applying it to each frame. Another application method is a method of applying the maximum delay time to each of the frames subsequent to the announcement of the information of the maximum delay time, i.e., to the frame for which the maximum delay time is indicated and to each of the frames temporally subsequent to that frame. The output, input, application, etc. of the maximum delay time in these methods will be specifically detailed later.

[0055]

The processing corresponding to the video encoding method executed in the foregoing video encoding apparatus 1 can be substantialized by the video encoding program for letting a computer to execute the video coding. The processing corresponding to the video decoding method executed in the video decoding apparatus 2 can be substantialized by the video decoding program for letting a computer to execute the video decoding.

[0056]

For example, the video encoding apparatus 1 can be constructed of a CPU connected to a ROM storing software programs necessary for respective operations of the video coding and a RAM temporarily saving data during execution of a program. In this configuration, the

video encoding apparatus 1 can be substantialized by letting the CPU to execute the predetermined video encoding program.

[0057]

Similarly, the video decoding apparatus 2 can be constructed of a CPU connected to a ROM storing software programs necessary for respective operations of the video decoding and a RAM temporarily saving data during execution of a program. In this configuration, the video decoding apparatus 2 can be substantialized by letting the CPU to execute the predetermined video decoding program.

[0058]

The above-stated program for letting the CPU to execute the processes for video encoding or for video decoding can be distributed in a form in which it is recorded in a computer-readable recording medium. Such recording media include, for example, magnetic media such as hard disks and floppy disks, optical media such as CD-ROM and DVD-ROM, magnetooptical media such as floptical disks, or hardware devices, for example, such as RAM, ROM, and semiconductor nonvolatile memories, specially mounted to execute or store program commands.

[0059]

The video encoding apparatus, the video decoding apparatus, the video processing system provided therewith shown in Fig. 1, and the video encoding method and video decoding method corresponding thereto will be described with specific embodiments. The description hereinafter will be based on the presumption that the encoding and decoding operations of motion video are implemented based on H.26L,

and parts not specifically described about the operation in video encoding will be pursuant to the operation in H.26L. It is, however, noted that the present invention is not limited to H.26L.

[0060]

(First Embodiment)

First, the first embodiment of the present invention will be described. The present embodiment will describe an embodied form of encoding at a fixed frame rate. In the encoding according to the present embodiment, the maximum number of reference frames used for backward prediction is first determined, the maximum delay time is calculated thereafter from this maximum number of reference frames and the frame rate used in encoding, and the maximum delay time is then outputted. In the decoding according to the present embodiment, on the occasion of decoding a backward-prediction-nonassociated frame, an output time of a decoded image thereof is delayed by the input maximum delay time. The delay time for the output time is uniformly applied to every frame thereafter, so as to prevent the output time interval between the decoded image of the backward-prediction-nonassociated frame and the decoded image of the backward predicted frame from deviating from the original interval.

[0061]

In the encoding, since the upper bound of the number of reference frames used is preliminarily determined, the maximum number of reference frames used for backward prediction is first determined within the range not exceeding the upper bound. Then, based on the frame rate used for encoding, which is also preliminarily

determined, the maximum delay time is calculated as a time interval of one frame or two or more frames according to the maximum number of reference frames used for backward prediction.

[0062]

Fig. 2 is a diagram showing an example of encoding of a frame in execution of bidirectional prediction. Here this Fig. 2 shows the example in which reference frames used for the current frame F2 are two temporally previous frames F0, F1 before the current frame F2 and two temporally subsequent frames F3, F4 after the current frame F2.

[0063]

In the case where the maximum number of reference frames used for backward prediction is 2 and where the frame rate is 15 frames/second, as shown in Fig. 2, the time interval of one frame is 1/15 second. In this case, therefore, the maximum delay time is $2 \times (1/15) = 2/15$ second.

[0064]

In the encoding operation, encoding of each frame hereinafter is controlled so as not to carry out backward prediction requiring a delay time over the maximum delay time. Specifically, a sequence of encoding of frames is controlled so that any reference frame used in backward prediction, i.e., any temporally subsequent frame after the current frame is not encoded and outputted prior to the current frame over the maximum number of reference frames used in backward prediction.

[0065]

It is assumed in the present embodiment that a syntax for

transmitting the maximum delay time is added to the encoded data syntax in H.26L, in order to implement the output of the maximum delay time in the encoding and the input of the maximum delay time in the decoding. In this example the new syntax is added into the Sequence Parameter Set being a syntax for transmitting the information to be applied to the entire encoded data.

[0066]

The parameter init_output_delay is defined as a syntax for carrying the maximum delay time. It is assumed here that the parameter init_output_delay uses the same time unit used in the other syntaxes indicating the time in H.26L and that it indicates the maximum delay time in the time unit of 90 kHz. A numeral indicated in the time unit is encoded and transmitted by a 32-bit unsigned fixed-length code. For example, where the maximum delay time is 2/15 second as described above, init_output_delay is $(2/15) \times 90000 = 12000$.

[0067]

In the decoding operation, the maximum delay time carried by init_output_delay is decoded, and an output time of a decoded image is delayed using it.

[0068]

Fig. 3 is a figure including diagrams showing (a) decoding and (b) output of the frames in the case of the bidirectional prediction shown in Fig. 2. It is assumed in the decoding operation that the encoded data of the frames is decoded in the order necessary for execution of the interframe prediction, the intervals thereof are constant time intervals according to the frame rate, and the time necessary for the decoding

operation is negligible for each frame, regardless of whether interframe prediction is applied and regardless of the directions of interframe prediction. In this case, the maximum delay time necessary for execution of the backward prediction in the backward predicted frame is equal to a time interval of a frame or frames according to the maximum number of reference frames used for the backward prediction. This time is carried as a maximum delay time by init_output_delay. Accordingly, for outputting a decoded image, an output time thereof is delayed by the maximum delay time.

[0069]

In practice, the decoding intervals of the respective frames are not constant, and can vary depending upon such factors as variation in encoding bits of the frames. The time necessary for the decoding operation of each frame can also vary according to whether the frame is a backward predicted frame or according to encoding bits of each frame.

[0070]

For delaying the output time, therefore, the reference is set at the time when the decoded image is obtained for the backward-prediction-nonassociated frame F0 without delay due to backward prediction and without reversal of orders of decoding times and output times with respect to any other frame, as shown in Fig. 3. Namely, a time obtained by delaying the time when the decoded image is obtained, by the maximum delay time announced by init_output_delay is defined as a time equal to the output time correlated with this decoded image, and is used as a reference time in output of decoded images. The decoded images F1-F4 thereafter are outputted

when this reference time agrees with a time equal to an output time correlated with each decoded image.

[0071]

For example, where the maximum delay time is 2/15 second as described above, a time at a delay of 2/15 second from the time when the decoded image is obtained for the backward-prediction-nonassociated frame, is defined as a time equal to the output time correlated with this decoded image and is used as a reference time in output of decoded images thereafter.

[0072]

According to the circumstances, conceivably, the maximum delay time is not announced on purpose, in order to simplify the encoding or decoding operation. For such cases, the syntax for announcing the maximum delay time may be arranged to be omissible on the presumption that a flag to indicate the presence or absence of the syntax is transmitted prior to the syntax for transmitting the maximum delay time.

[0073]

In the case where the announcement of the maximum delay time is omitted, the encoding operation may be preliminarily stipulated, for example, so as not to use the backward prediction or so that the number of reference frames used in backward prediction can be optionally altered within the range not exceeding the upper bound of the number of reference frames.

[0074]

The decoding operation may be configured to perform in

conformity with the stipulation in the encoding operation, for example, when backward prediction is not applied, there occurs no delay necessary for execution of backward prediction; or, the decoding operation may also be configured so that the number of reference frames used in backward prediction can be optionally altered within the range not exceeding the upper bound of the number of reference frames, i.e., the delay time can vary large. In this case, the decoding operation may be configured to always perform processing assuming an expected maximum delay time, or the decoding operation may be configured to allow variation of output time intervals of decoded images and perform simplified processing without consideration to the delay time of each frame.

[0075]

The present embodiment was described on the assumption that the operations were implemented based on H.26L, but it is noted that the video encoding methods to which the present invention can be applied are not limited to H.26L and that the present invention can be applied to various video encoding methods using the backward interframe prediction.

[0076]

In the present embodiment, the syntax by fixed-length codes was added as a syntax for transmitting the maximum delay time into the Sequence Parameter Set, but it is noted that the codes and syntax for transmitting it, or the time unit for expressing the maximum delay time are not limited to these, of course. The fixed-length codes may be replaced by variable-length codes, and the maximum delay time can be

transmitted by any of various syntaxes that can convey information to be applied to the entire encoded data.

[0077]

For example, in H.26L, a syntax may be added into a Supplemental Enhancement Information Message. In a case using another video encoding method, the maximum delay time may be transmitted by a syntax for transmitting the information to be applied to the entire encoded data in the pertinent encoding method. In another case, the maximum delay time may also be transmitted outside the encoded data in the video encoding method as in ITU-T Recommendation H.245 used for conveying control information in communication using H.263.

[0078]

(Second Embodiment)

The second embodiment of the present invention will be described below. The present embodiment will describe an embodied form of encoding at variable frame rates. The operations in the encoding and decoding according to the present embodiment are basically much the same as in the first embodiment. Since the present embodiment uses the variable frame rates, it involves an operation at low frame rates to avoid execution of the backward prediction requiring the delay time over the preliminarily calculated maximum delay time, in addition to the operation in encoding in the first embodiment, so as to prevent the output time interval between the decoded image of the backward-prediction-nonassociated frame and the decoded image of the backward predicted frame from deviating from the original interval

even with variation of frame rates.

[0079]

Since in the encoding operation the upper bound of the number of reference frames is preliminarily determined, the maximum number of reference frames used for backward prediction is first determined within the range not exceeding the upper bound. Then the maximum frame time interval is determined based on a target frame rate preliminarily determined in control of encoding bit rates, and the maximum delay time is calculated as a time interval of one frame or two or more frames according to the maximum number of reference frames used in backward prediction and the maximum frame time interval.

[0080]

In the encoding operation, encoding of each frame thereafter is controlled so as to avoid the backward prediction requiring the delay time beyond the maximum delay time. Specifically, the order of encoding of frames is controlled so as to prevent any reference frame used in backward prediction, i.e., any temporally subsequent frame after the current frame, that goes beyond the maximum number of reference frames used in backward prediction, from being encoded and outputted prior to the current frame.

[0081]

In addition, when the encoding frame rate becomes temporarily small because of control of encoding bit rates, so as to make the frame time interval in that case larger than the maximum frame time interval, encoding of each frame is controlled so as not to apply backward prediction to encoding of the frame there.

[0082]

The present embodiment is substantially identical to the first embodiment in that the maximum delay time is outputted in the encoding, in that the syntax init_output_delay to transmit the maximum delay time is added to the encoded data syntax in order to effect input thereof in the decoding, and in the definition of the syntax.

[0083]

In the present embodiment, the decoding operation is arranged to decode the maximum delay time announced by init_output_delay and delay the output time of the decoded image by use of it. This processing is also the same as in the first embodiment.

[0084]

(Third Embodiment)

The third embodiment of the present invention will be described below. The present embodiment will describe an embodied form in which the maximum delay time is optionally announced for each frame and is thus flexibly changeable. The operations in the encoding and decoding according to the present embodiment are basically similar to those in the first embodiment or the second embodiment.

[0085]

In the present embodiment, the syntax init_output_delay to transmit the maximum delay time, which was defined in the first embodiment, is arranged to be added into the Picture Parameter Set being a syntax to carry the information applied to each frame instead of the syntax to carry the information applied to the entire encoded data. The syntax init_output_delay herein is configured to indicate the

maximum delay time in the time unit of 90 kHz, as in the case of the first embodiment, and a numeral expressed in the time unit is encoded and transmitted by a 32-bit unsigned fixed-length code.

[0086]

The present embodiment is much the same as the first embodiment, as to the calculation of the maximum delay time in encoding and as to the delay of the output time of the decoded image by use of the maximum delay time in decoding.

[0087]

Since the maximum delay time defines the reference time in output of decoded images from the time when the decoded image of the backward-prediction-nonassociated frame is acquired, it is enough to transmit the maximum delay time only for the backward-prediction-nonassociated frame. It is therefore possible to employ, for example, a configuration wherein the syntax for transmitting the maximum delay time is arranged to be omissible on the presumption that a flag indicating the presence or absence of the syntax is transmitted prior thereto. The syntax may also be arranged to be optionally omitted for the backward-prediction-nonassociated frame, provided that the maximum delay time transmitted before is applied in that case where the maximum delay time is not transmitted.

[0088]

The syntax for each frame in the present embodiment may also be used simultaneously with the syntax for the entire encoded data as defined in the first embodiment. In this case, the syntax for each frame is omissible, provided that a flag indicating the presence or absence of

the syntax is transmitted prior thereto as described above. The maximum delay time transmitted in the syntax for the entire encoded data is continuously applied before the maximum delay time is transmitted in the syntax for each frame. After it is updated by the syntax for each frame, the time delayed based thereon is used as a reference time in output of every decoded image thereafter.

[0089]

The present embodiment was described on the assumption that it was substantialized based on H.26L, but it is noted that the video encoding methods to which the present invention can be applied are not limited to H.26L and that the present invention can be applied to various video encoding methods using the backward interframe prediction.

[0090]

In the present embodiment the syntax for transmitting the maximum delay time was the syntax by fixed-length codes added into the Picture Parameter Set, and it is a matter of course that the codes and syntax for transmitting it, or the time unit for expressing the maximum delay time are not limited to these, of course. The fixed-length codes can be replaced by variable-length codes, and the maximum delay time can be announced in any of various syntaxes capable of announcing the information to be applied to each frame.

[0091]

For example, the syntax may be added into a Supplemental Enhancement Information Message in H.26L. When another video encoding method is applied, it is possible to use a syntax for announcing information to be applied to each frame in the pertinent encoding

method. In addition, the information may also be announced outside the encoded data in the video encoding method as in ITU-T Recommendation H.245 used for announcement of control information in communication using H.263.

[0092]

[Effects of the Invention]

The video encoding method, video decoding method, video encoding apparatus, video decoding apparatus, video processing system, video encoding program, and video decoding program according to the present invention provide the following effect, as detailed above. Namely, when a moving picture consisting of a series of frames is encoded by the backward interframe prediction to be outputted, it becomes feasible to achieve output of decoded images at appropriate time intervals when employing the backward interframe prediction, by the video encoding method, encoding apparatus, and encoding program configured to output the maximum delay time due to the backward prediction, the video decoding method, decoding apparatus, and decoding program configured to effect input of the maximum delay time, and the video processing system using them.

[Brief Description of the Drawings]

[Fig. 1] A block diagram showing the schematic structure of the video encoding apparatus, video decoding apparatus, and video processing system.

[Fig. 2] A diagram showing an example of encoding of frames in the case of the bidirectional prediction being carried out.

[Fig. 3] A figure including diagrams showing (a) decoding and

(b) output of frames in the case of the bidirectional prediction shown in Fig. 2 being carried out.

[Fig. 4] A diagram showing encoding of frames in the case of the bidirectional prediction being carried out.

[Fig. 5] A figure including diagrams showing (a) decoding and (b) output of frames in the case of the bidirectional prediction shown in Fig. 4 being carried out.

[Fig. 6] A diagram showing encoding of frames in the case of the bidirectional prediction being carried out.

[Fig. 7] A figure including diagrams showing (a) decoding and (b) output of frames in the case of the bidirectional prediction shown in Fig. 6 being carried out.

[Fig. 8] A figure including diagrams showing (a) decoding, (b) output, and (c) delayed output of frames in the case of the bidirectional prediction being carried out.

[Fig. 9] A figure including diagrams showing encoding of frames in the case of the bidirectional prediction being carried out.

[Explanation of Reference Numerals]

1 - video encoding apparatus, 2 - video decoding apparatus,

D0 - video data, D1 - encoded data, D2 - decoded video data, F0 - F4 - frame.

[Document Name] Abstract

[Abstract]

[Object] To provide a video encoding method, video decoding method, video encoding apparatus, video decoding apparatus, video processing system, video encoding program, and video decoding program capable of achieving output of decoded images at appropriate time intervals in application of backward interframe prediction.

[Means of Solution] A video processing system is provided with video encoding apparatus 1 and video decoding apparatus 2. The encoding apparatus 1 outputs a maximum delay time that is incurred by backward prediction, in addition to encoded data D1 resulting from encoding of video data D0. The decoding apparatus 2 effects input of the maximum delay time that is incurred by backward prediction, in addition to encoded data D1 from the encoding apparatus 1. Then, the encoded data D1 is decoded with reference to the input maximum delay time to generate motion video data D2.

[Selected Drawing] Fig. 1

Fig.1

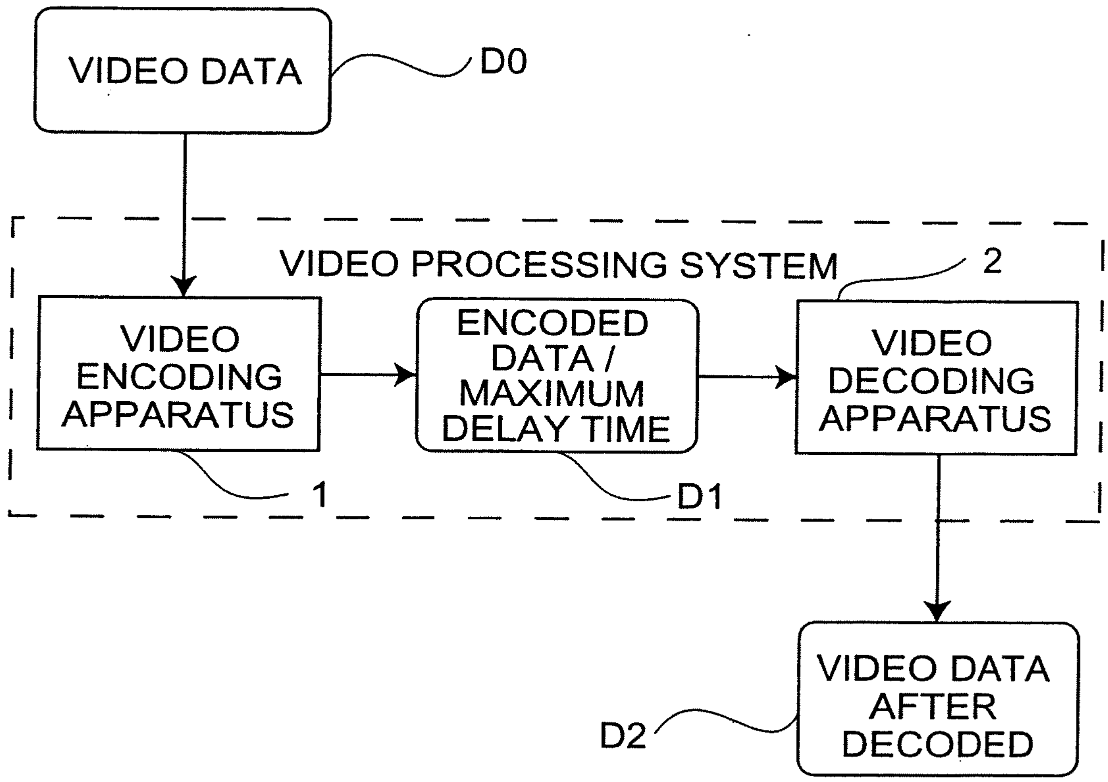

VIDEO DATA
D0
VIDEO PROCESSING SYSTEM
2
VIDEO ENCODING APPARATUS
ENCODED DATA / MAXIMUM DELAY TIME
VIDEO DECODING APPARATUS
1
D1
VIDEO DATA AFTER DECODED
D2

Fig.2

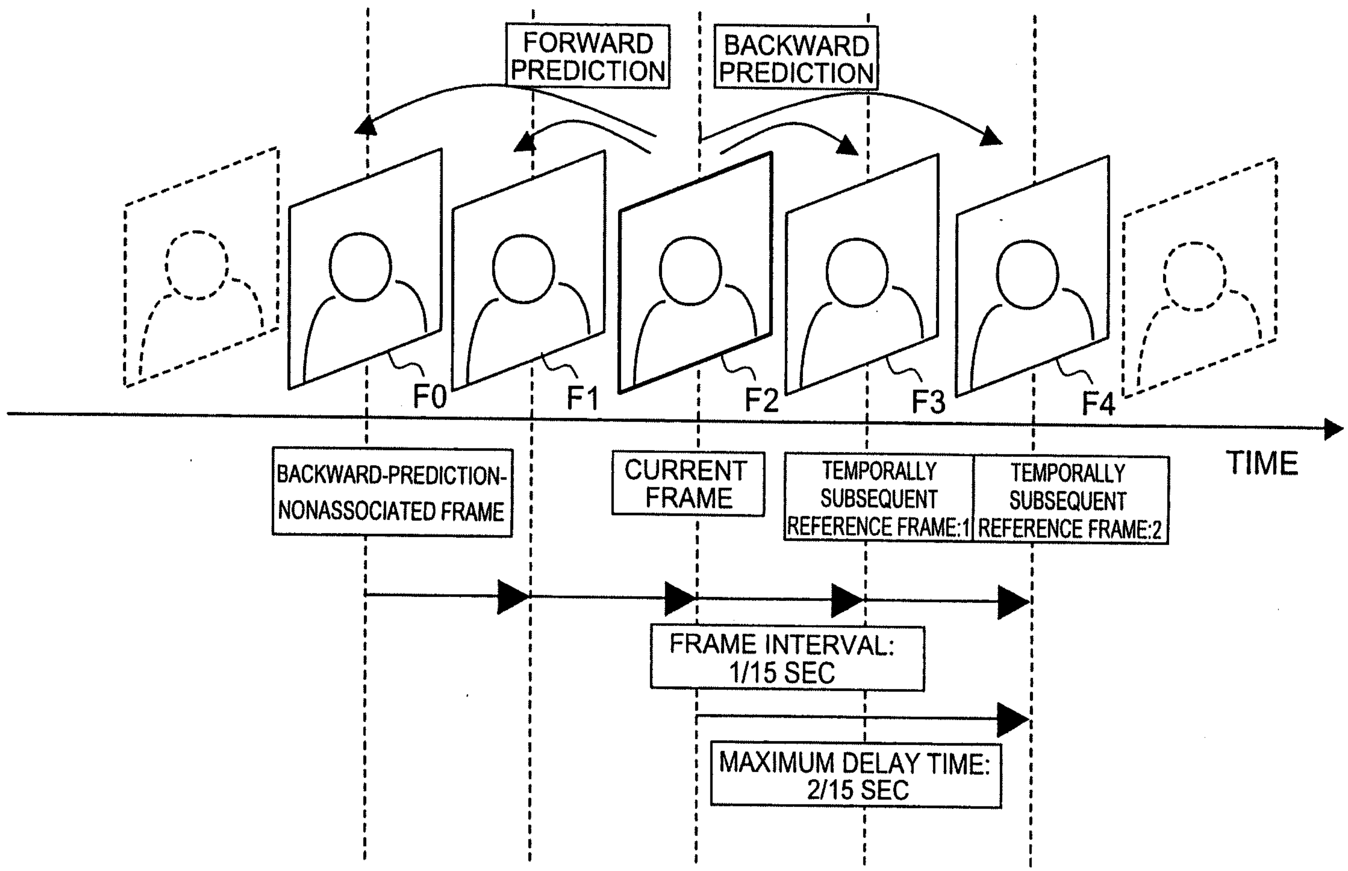
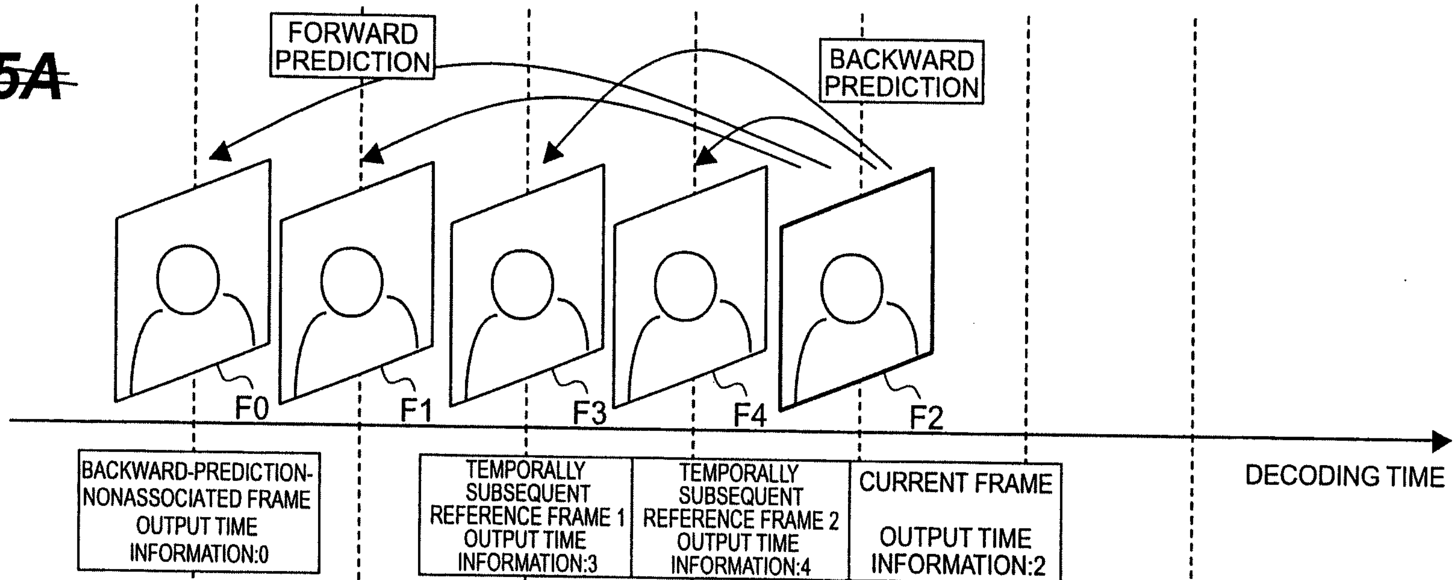

FORWARD PREDICTION
BACKWARD PREDICTION
F0
F1
F2
F3
F4
TIME
BACKWARD-PREDICTION-NONASSOCIATED FRAME
CURRENT FRAME
TEMPORALLY SUBSEQUENT REFERENCE FRAME:1
TEMPORALLY SUBSEQUENT REFERENCE FRAME:2
FRAME INTERVAL: 1/15 SEC
MAXIMUM DELAY TIME: 2/15 SEC

Fig. 3

~~Fig.5A~~
(a)




~~Fig.5B~~
(b)

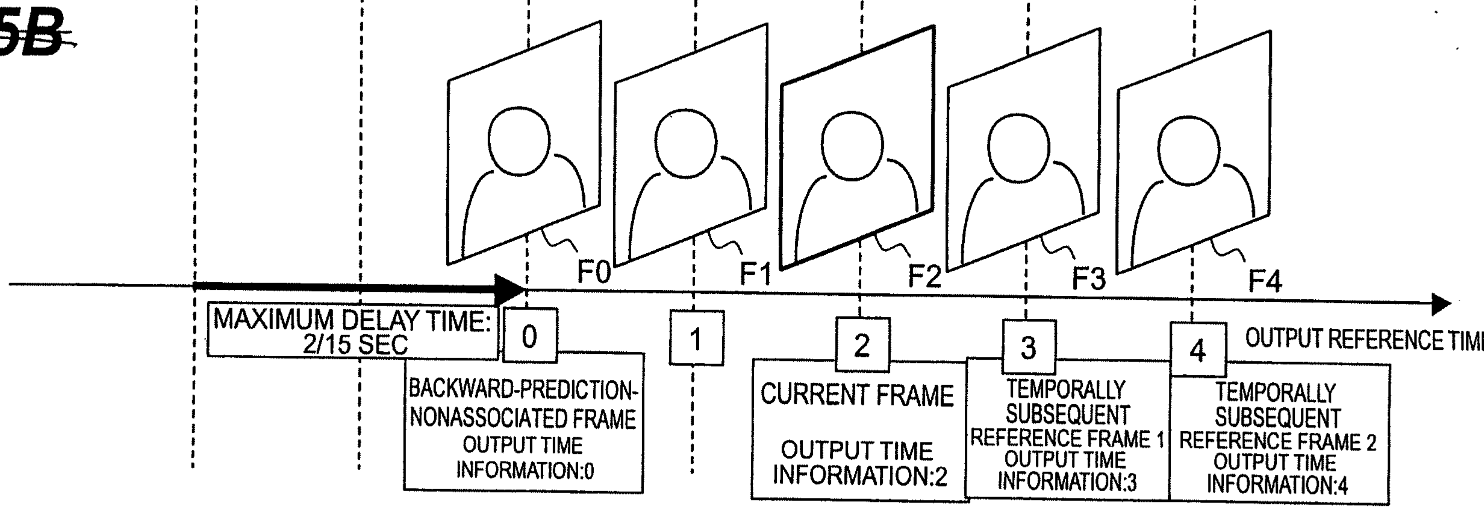


Fig. 4

~~Fig.6~~

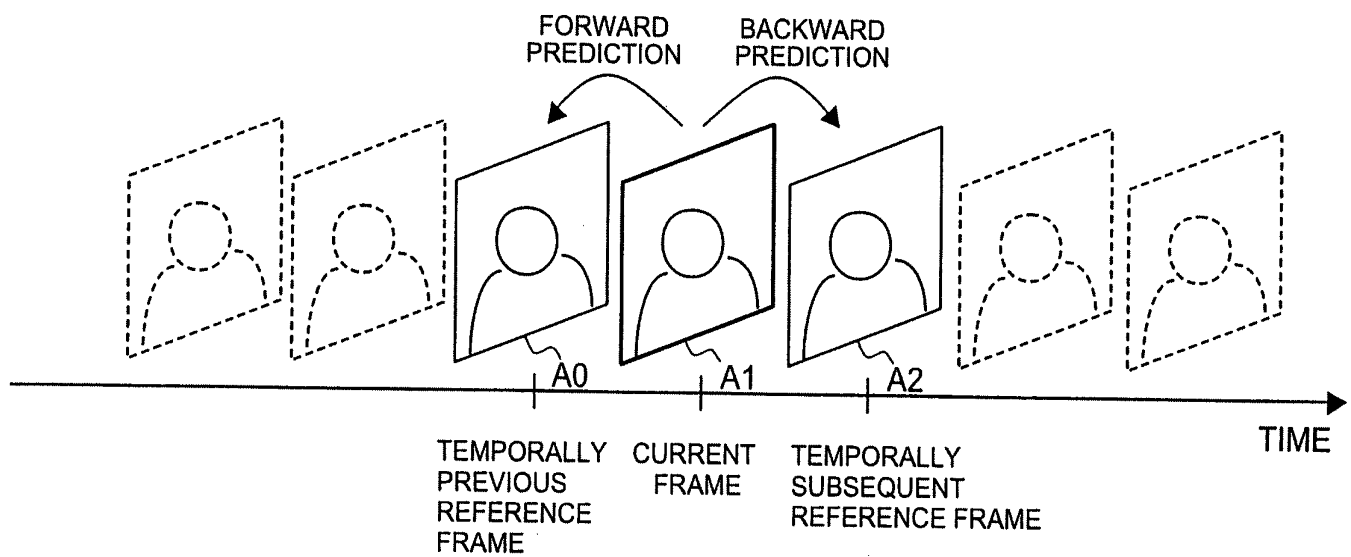
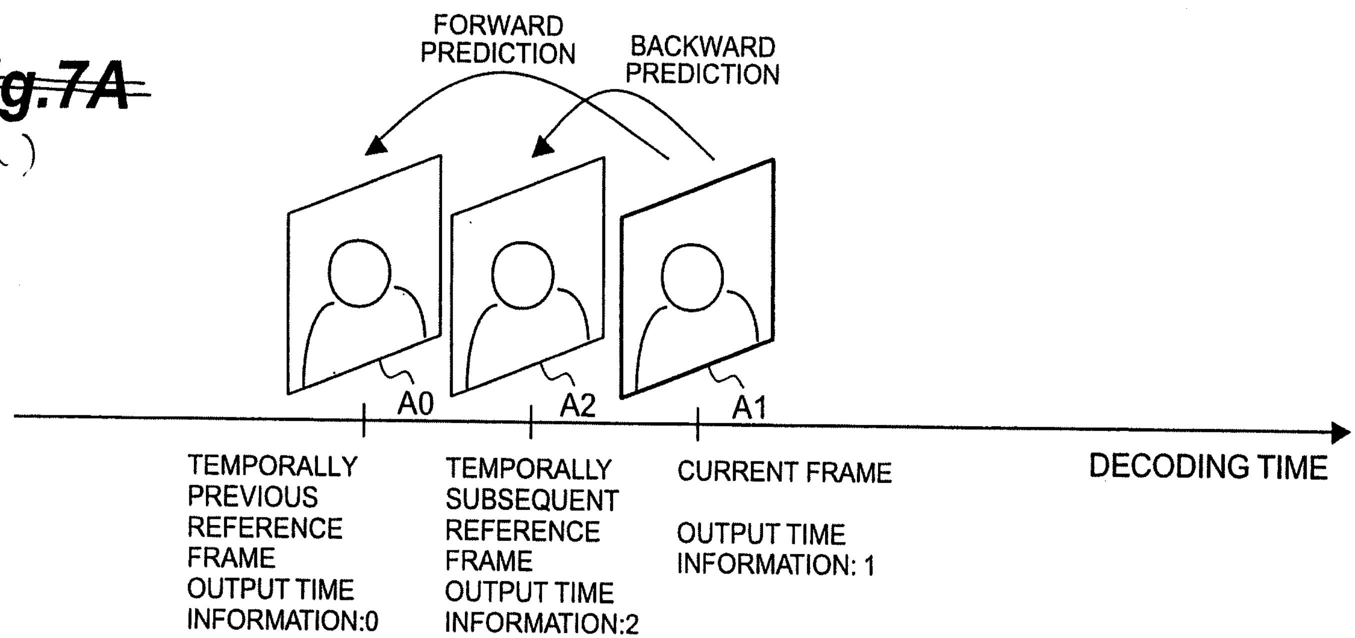

FORWARD PREDICTION
BACKWARD PREDICTION
A0
A1
A2
TIME
TEMPORALLY PREVIOUS REFERENCE FRAME
CURRENT FRAME
TEMPORALLY SUBSEQUENT REFERENCE FRAME

Fig. 5

~~Fig.7A~~
(a)




~~Fig.7B~~
(b)

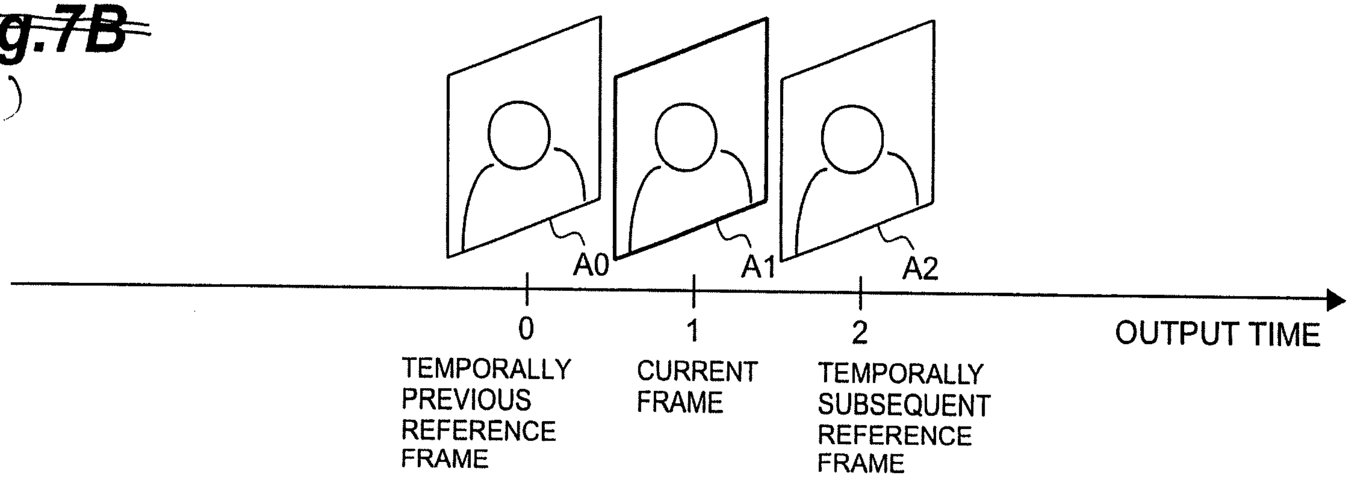


Fig. 6

~~Fig.8~~

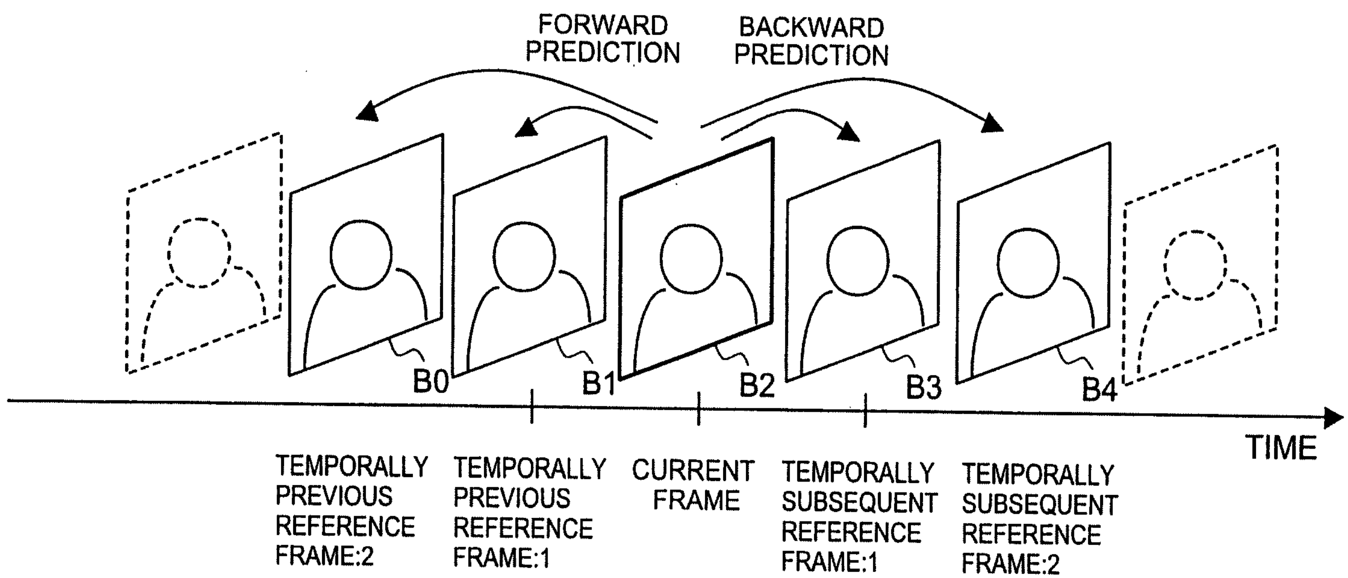
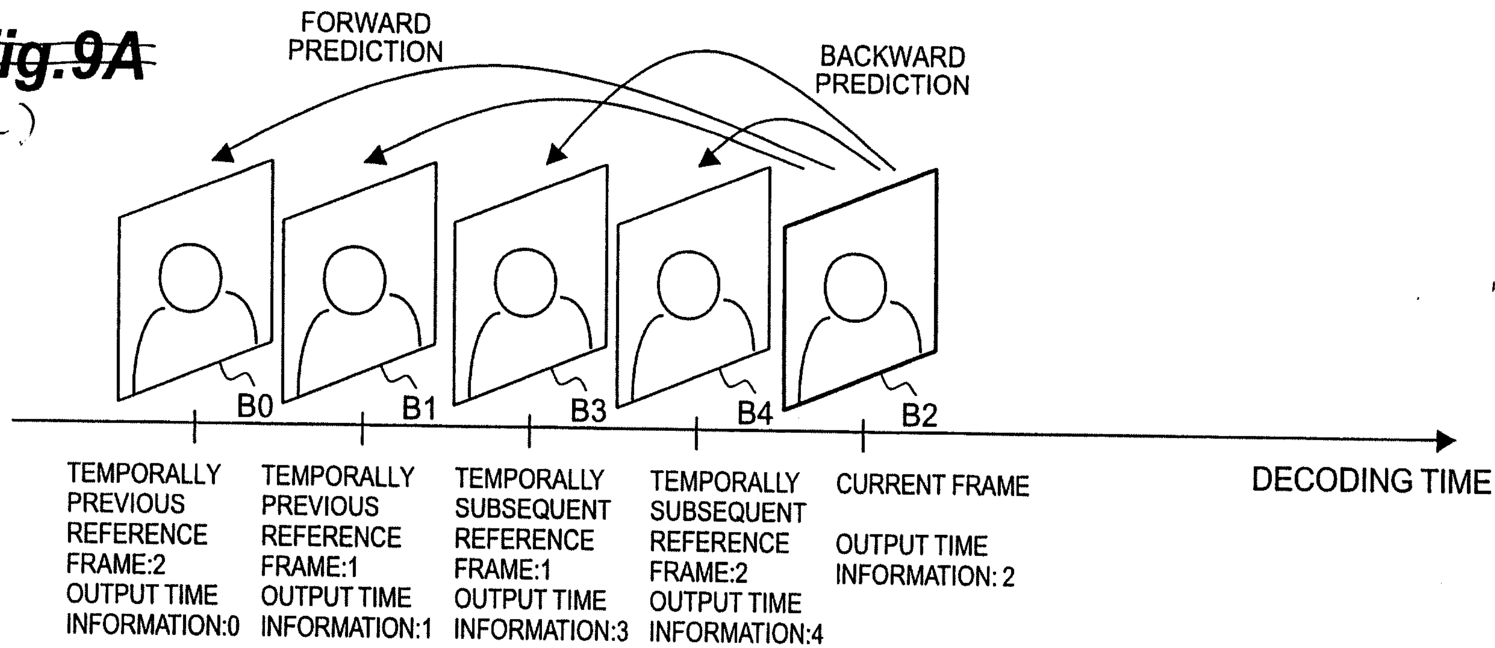

FORWARD PREDICTION
BACKWARD PREDICTION
B0
B1
B2
B3
B4
TIME
TEMPORALLY PREVIOUS REFERENCE FRAME:2
TEMPORALLY PREVIOUS REFERENCE FRAME:1
CURRENT FRAME
TEMPORALLY SUBSEQUENT REFERENCE FRAME:1
TEMPORALLY SUBSEQUENT REFERENCE FRAME:2

Fig. 7

~~Fig.9A~~
(a)




~~Fig.9B~~
(b)

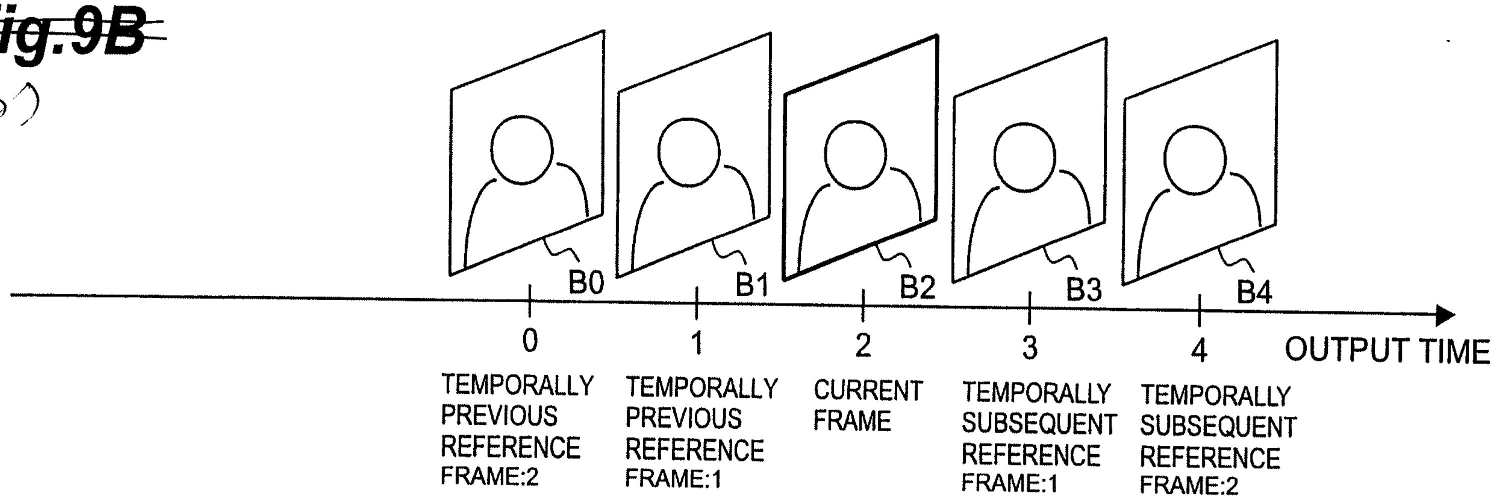
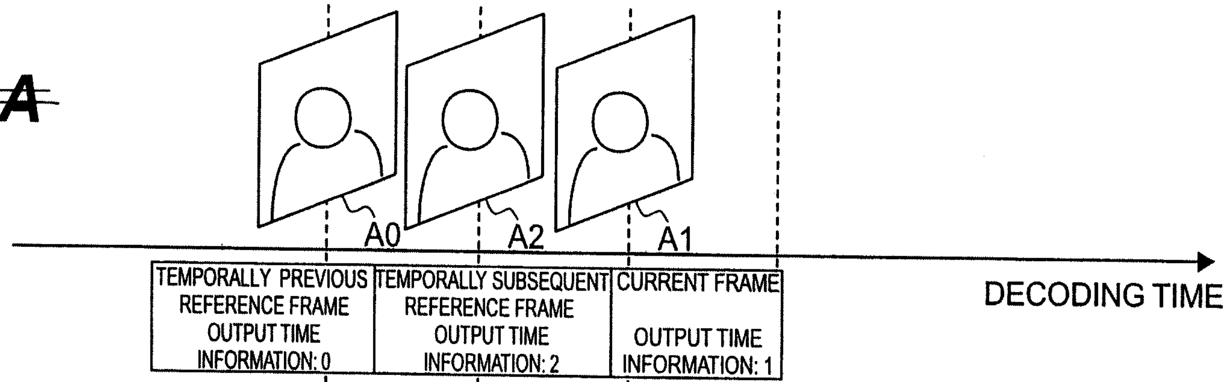
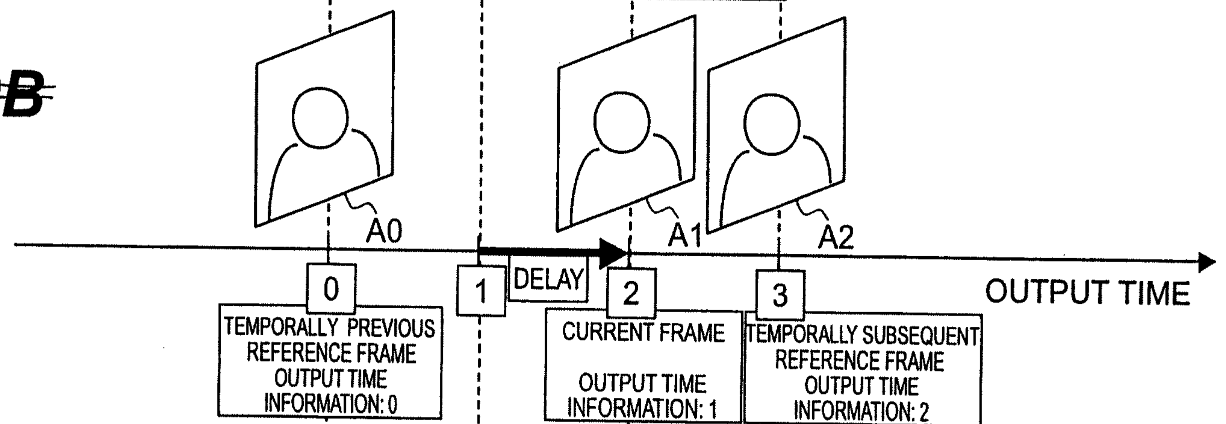


Fig. 8

~~Fig.10A~~ (a)




~~Fig.10B~~ (b)




~~Fig.10C~~ (c)

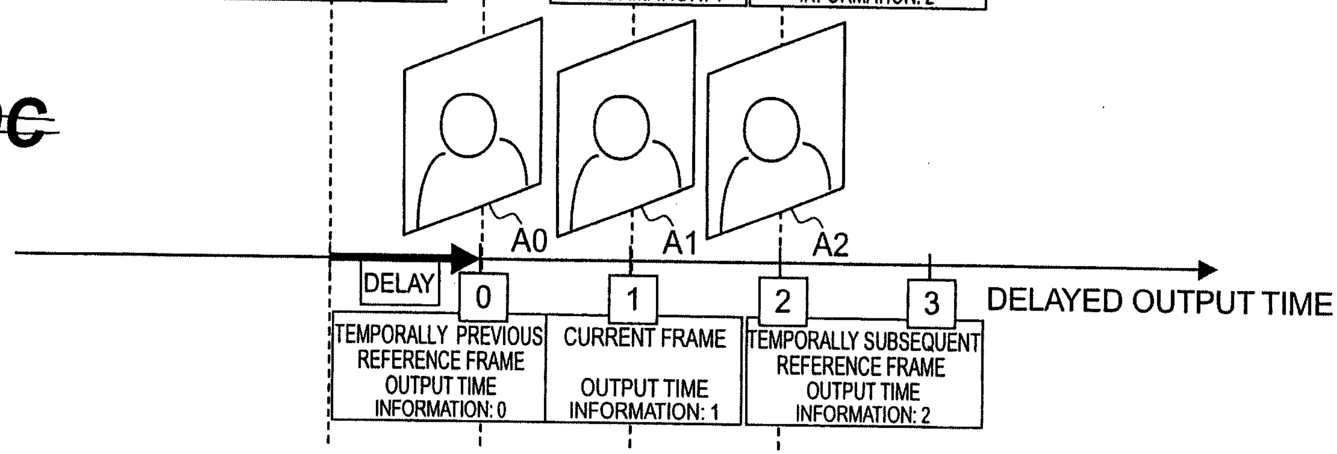
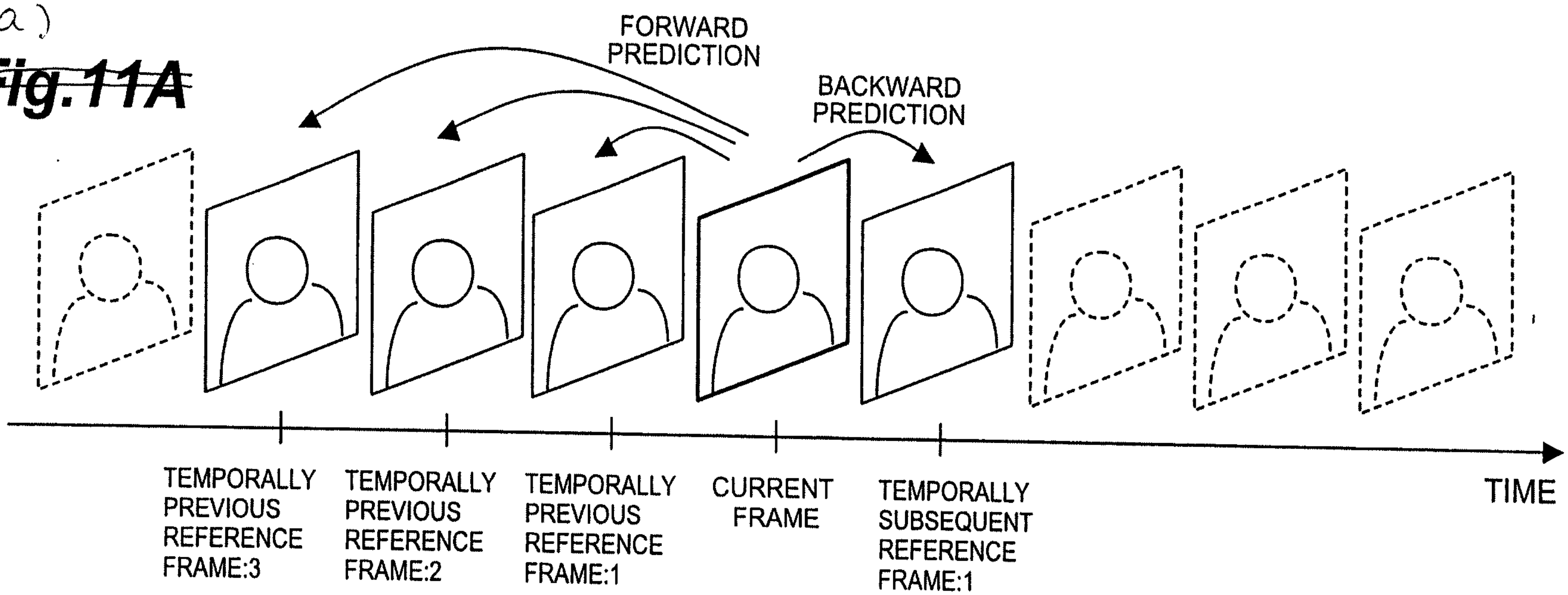


Fig. 9

(a)

~~Fig.11A~~




(b)

~~Fig.11B~~